AD-A191 559

OPTIMIZING FEEDBACK UTILIZATION

IN MOTOR SKILL TRAINING

Richard A. Schmidt and Diane C. Shapiro
University of California at Los Angeles

for

DTIC
SELECTED
FEB 2 5 1988
D

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

February 1988

88 2 24 088

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

University of California at Los Angeles

Technical review by

Dan Ragland

Accesion For

NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	

By

Distribution /

Availability Codes

| Dist | Avail and/or Special |
| A-1 | |

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 88-05	2. GOVT ACCESSION NO. ADA191559	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Optimizing Feedback Utilization in Motor Skills Training		5. TYPE OF REPORT & PERIOD COVERED Interim Report June 85 - June 86
		6. PERFORMING ORG. REPORT NUMBER - -
7. AUTHOR(s) Richard A. Schmidt and Diane C. Shapiro		8. CONTRACT OR GRANT NUMBER(s) MDA903-85-K-0225
9. PERFORMING ORGANIZATION NAME AND ADDRESS Motor Control Laboratory, Dept. of Kinesiology, University of California at Los Angeles, Los Angeles, CA 90024		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE February 1988
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(*If different from Controlling Office*) - -		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE - -
16. DISTRIBUTION STATEMENT *(of this Report)* Approved for public release, distribution unlimited.		
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)* - -		
18. SUPPLEMENTARY NOTES Judith Orasanu, contracting officer's representative		
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)* Motor Learning Transfer Knowledge of Results Retention Error Estimation Feedback Motor Skills Training		

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This research note deals with the acquisition of motor skills, specifically with the optimal use of feedback on goal achievement (termed knowledge of results, KR) for the maximization of learning and retention. A great deal of evidence suggests that enhancing the amount or quality of KR can improve performance in a session where KR is present, but limited evidence suggests that this might not be effective on a delayed criterion test when KR is removed -- a typical goal of many Army training settings. Two variations of KR were

(OVER)

DD _{1 JAN 73} FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE

ARI Research Note 88-05

20. <u>Abstract</u> (continued)

studied: relative frequency - the proportion of trials receiving KR, and summary KR - where KR is given about an entire set of trials. In each case, alterations in KR which degraded training performance (relative to a condition with KR on each trial) actually produced <u>enhanced</u> performance on on a delayed no-KR retention test. A third paradigm examined the learning of error-detection capabilities as a basis for these effects. Overall, six experiments suggest that enhancing KR in acquisition may generate KR overreliance, preventing the learning of important features of the task which are critical for retention of performance when feedback is removed or degraded (e.g. in marksmanship). Our results have implications for improved feedback in training.

OPTIMIZING FEEDBACK UTILIZATION

IN MOTOR SKILL TRAINING

Richard A. Schmidt and Diane C. Shapiro

1

SUMMARY

In the concern of training for skilled motor performance, it is generally recognized that feedback about movement proficiency--termed knowledge of results (KR)--is critical to efficient learning. But, while various experimental manipulations of KR in training can provide immediate benefits for performance, these benefits may disappear when learners are asked to perform on retention tests where KR is not provided, as in many Army training situations. In fact, the literature shows that some variations in KR may benefit performance in training, but may result in poorer performance in retention. This project examined a number of such variations in KR for learning and retaining simple discrete movement skills.

In one paradigm, we examined the relative frequency of KR, or the proportion of trials on which KR is provided. Whereas decreasing the relative frequency slows the rate of acquisition when KR is present, it results in superior retention. Another paradigm involves summary KR, in which KR about a set of trials is only provided after that set had been completed. Increasing the number of summarized trials decreases performance in acquisition, but greatly enhances retention. In both cases, providing degraded feedback information in training produced poorer performance in training, but superior retention. We hypothesize that providing KR on every trial produces a kind of reliance on it, preventing the subject from learning other important features of the movement, which results in poor performance when KR is withdrawn in a retention test. For example, subjects with frequent KR may not develop important capabilities to detect their own errors--which is not a problem during training when KR is being given, but becomes critical when the KR is later withdrawn. In a third paradigm, we ask (or do not ask) subjects to estimate their own errors in an attempt to experimentally strengthen or degrade these error-detection capabilities, and we find some support for the hypothesis that these processes are related to retention.

These results should have important implications for Army training, providing guidance about the scheduling of feedback in skills (such as riflery) where the important goal is the capability to perform in the future, often under degraded feedback conditions. More effective KR utilization could increase readiness and reduce the costs of, and needs for, retraining.

BACKGROUND AND RATIONALE

Scientists who have studied the acquisiton of motor skills have known for centuries that feedback to learners about their success in meeting some environmental goal is critical for efficient learning (see Adams, 1986, for an historical review), and the area has attracted considerable attention as a result. This feedback is usually operationalized in laboratory settings as knowledge of results (KR), defined as verbal(izable), post-response information about the score for the task being attempted (Schmidt, 1982, Chapter 13). The literature is remarkably consistent in showing that, during acquisition (i.e., a practice phase) with simple motor tasks, increasing the amount (or frequency) of KR, making it more precise, or in general giving more information to the learner, benefits performance and increases the rate of improvement over trials (Newell, 1974; Salmoni, Schmidt, & Walter, 1984).

It has been tempting to conclude on the basis of this evidence that increasing the amount of information learners receive is uniformly beneficial to learning. But it ignores the possibility--first recognized long ago (Guthrie, 1952; Hull, 1943; Tolman, 1929)--that the changes in performance seen in practice are not necessarily due to the relatively permanent changes in the capability for responding that we usually term "learning," but rather may be due to temporary, transient factors which could disappear as soon as the variation in KR is removed. Thus, when learners are returned to the task after a period of no-practice, they may show decrements in performance relative to levels achieved during practice. But more importantly, one variation in KR which improves (relative to another variation) performance in acquisition, may actually turn out to be less effective for performance in a retention test, especially if the retention test involves a "degraded" feedback condition as shooting usually does in Army settings. One solution to this research problem is to use delayed retention tests in which the KR conditions are equated (e.g., a no-KR retention test); here, the temporary (performance) effects of the KR manipulations are argued to have dissipated, leaving the relatively permanent (learning) effects as evidence for the effectiveness of the KR variation. The problem is that almost all the experiments on the role of feedback in motor learning have failed to consider this learning-performance distinction. As a result, we are left with the question of whether the large effects of KR variations--found during an acquisition phase--are learning effects, merely performance effects, or perhaps some combination of the two.

But, in our review (Salmoni et al., 1984), we did locate a few KR investigations which have used no-KR retention tests, and the findings here are generally quite surprising. In two lines of work, variations in KR which degraded performance in acquisition (relative to the "usual" condition with KR after each trial) were shown to improve performance in a no-KR retention test performed a few days later. One variable was relative frequency of KR, or the proportion of trials on which KR was given (Ho & Shea, 1978; Johnson, Wicks, & Ben-Sira, 1981), where low (e.g., 10%) relative

frequencies of KR in acquisition degraded accuracy relative to a 100% condition, yet resulted in markedly more accurate performance on a delayed no-KR retention test. Another example involved "summary KR," in which a summary of a set of trials (e.g., 20) was provided only after that set had been completed (Lavery, 1962; Lavery & Suddon, 1962). During acquisition, summary KR degraded performance relative to a condition in which KR was given after each trial; but on a delayed no-KR retention test, groups which had received summary KR performed considerably more accurately than groups who had received KR after each trial. These results show that at least some variations of KR which are beneficial for performance in acquisition are detrimental to the development of a long-term capability to perform. Of course, these effects of KR could be due to (a) more effective learning during acquisition (b) increased resistance to forgetting during the retention interval, or both.

VARIATIONS OF KR TO MAXIMIZE LEARNING AND RETENTION
Rationale and Potential Applications

As just described, those variations in feedback which enhance performance in acquisition are not necessarily those which will produce optimum performance in delayed retention tests. This is a particularly important problem in Army settings such as marksmanship training. First, feedback is an important component of this training, and is easily manipulated by instructors or incorporated into simulation devices. Second, marksmanship is a skill for which the criterion is not necessarily success in a training session, but rather is success in a delayed retention test (e.g., on a battlefield). Further, these conditions often involve degraded feedback, often with poor lighting or dusty, smokey conditions, so that the soldier cannot determine very easily whether or not a target has been hit. The laboratory experiments described above, where success in training can be evaluated in terms of performance on a delayed no-KR retention test, share many of the same features as this kind of Army training. Further, they suggest the possibility that efforts to provide high-fidelity KR and feedback in Army training could be overdone; this might allow the soldier to be very proficient in training because of various temporary phenomena, but might render him relatively ineffective on an important criterion test to be performed later.

Problems Investigated

But before such ideas can be accepted, various problems required attention, and our Contract was concerned with a number of these. First, many of the experiments in the literature suffer from various experimental design difficulties and have small amounts of practice. A first step was to repeat these results with more effective procedures. Second, most of the earlier KR experiments were done with very simple tasks. Thus another goal, to be done primarily in the second year, was to study these phenomena with more complex skills which would be more applicable to motor skills learning in general and to Army training situations in particular. A next phase was to investigate

these phenomena further, perhaps to determine the limits under which these variations in KR would be effective, and also to determine optimal levels for training. An additional goal involved tests of various hypotheses--based on information-processing perspectives--to explain the effects produced.

In the following sections, we outline the results of eight experiments completed in the first year of our project. The specific rationale for each of the studies, together with the important theoretical and practical implications, are described separately in each section.

EXPERIMENTAL RESULTS
Relative Frequency Effects (Paradigm A)[1]

Background

In this aspect of the project, we examined questions of the role of relative and absolute frequency of KR, begun some years ago by Bilodeau and Bilodeau (1958). In this work, absolute frequency of KR refers to the number of trials in a sequence for which KR is provided, whereas relative frequency of KR refers to the proportion of trials for which KR was provided; thus, relative frequency is simply the absolute frequency divided by the number of trials in acquisition. Earlier, Bilodeau and Bilodeau had shown that, for simple linear positioning tasks, the absolute frequency of KR was critical for performance in an acquisition phase (only), whereas relative frequency was not. No retention tests without KR were used, so it was not clear as to whether these effects were relatively permanent or only temporary in nature.

Ho and Shea (1978) and Johnson et al. (1981) extended Bilodeau and Bilodeau's (1958) findings by adding no-KR retention tests, also using simple linear positioning tasks. In both experiments, absolute frequency was held constant at 10 KRs, and the relative frequency was varied by altering the number of total trials (and, hence, the number of no-KR trials provided between the KR trials). The absolute errors in acquisition (when KR was present and being manipulated) were inversely related to the relative frequency, with 100% relative frequency conditions showing least error (and larger rates of improvement) at a given point in practice, and 10% relative frequency conditions showing the most error. However, in no-KR transfer tests conducted on a subsequent day, the 10% group retained its performance relatively well, whereas the 100% group had suffered considerable decrements; and, the error on the retention tests were in direct relation to the relative frequency conditions in acquisition, with the 100% conditions having most error, and the 10% conditions having the least. Clearly, lower relative frequency aided learning or retention, contrary to the earlier findings viewed only in the acquisition phase (e.g., Bilodeau & Bilodeau, 1958).

However, these later experiments confounded the amount of original practice with relative

---

1. Here and elsewhere, we label the experiments in this paradigm with the letter A to distinguish them from experiments in our other paradigms.

frequency. Since it is well known that practice, per se, influences learning and retention, it was unclear as to whether these effects of relative frequency were due to the scheduling of KR or to the amount of practice.

Experiment 1A

Our first experiment varied relative frequency conditions while controlling for the amount of practice. We tested the hypothesis that reducing the amount of KR provided (via lower relative frequency) in acquisition (total practice constant) would improve performance on delayed no-KR transfer tests.

Procedures. Undergraduates at UCLA, 24 males and 24 females, received practice on a ballistic-timing task involving right-hand movement of a slide along a trackway. When the slide left a starting position at the right, a timer began to run, and it stopped when the subject moved through a switch at the left end of the trackway, 84 cm away, thereby measuring movement time. In addition, the subject was to move to a 5-cm zone 63 cm from the starting position, then reverse direction to another 5-cm zone 46 cm from the starting position, and then proceed through the final switch at the left. The goal was to make this two-reversal movement with a total movement time of exactly 550 msec on each trial. KR, when given, was the actual movement time (in milliseconds). The intertrial interval was 10 sec, and KR delay (when given) was 5 sec.

The experiment involved three treatment conditions, with 16 subjects (8 males, 8 females) each. Two groups had 102 trials in an acquisition phase, but they differed in relative frequency of KR--100% (KR on every trial, labeled 102/1) in one group, and 33% (KR on every third trial, labeled 102/3) in another group. This contrast tests our suggestion that less KR (with practice constant) would be more effective for retention. In addition, a third condition controlled for absolute frequency; this condition (labeled 34/1) had 100% relative frequency, but with the same absolute frequency (34) as the 102/3 group, and of course only 34 acquisition trials. If absolute frequency is the only determinant of learning, as Bilodeau and Bilodeau (1958) argued, then Group 34/1 and 102/3 (with the same absolute frequency) should be similar in retention. Subjects had no-KR retention tests after both 10 min and 2 days.

Error measures. Performance scores were blocked across groups of trials, and the usual error scores (Schutz & Roy, 1973; Schmidt, Chapter 3) computed within subjects. We routinely compute four measures: total variability (E), the within-subject SD about the target as a measure of overall accuracy; variable error (VE), the within-subject SD about the subject's own mean, a measure of movement consistency; constant error (CE), a measure of average directional error with respect to the target, and a measure of bias; and absolute constant error (|CE|), the absolute value of each subject's CE, a measure of bias sensitive to the possibility that different subjects' biases are in different directions, which could "cancel" when the usual CE measure is used. Our general method is to use total variability (E) as a measure of "overall" accuracy, and analyze its components variable error (VE) and absolute constant error (|CE|).
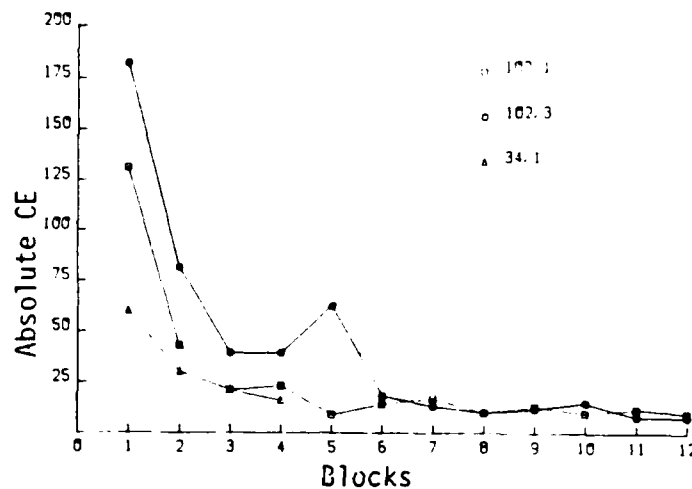
Figure 1. Mean absolute constant error (|CE|) over alternating 8- and 9-trial blocks in the acquisition phase for Experiment 1A; Group 102/1 had 102 trials with KR after each trial, Group 102/3 had 102 trials with KR after every third trial, and Group 34/1 had 34 trials with KR after each trial.

Results--acquisition phase. Figure 1 shows the average |CE| as a function of blocks of trials in the acquistion phase. The two groups with 100% relative frequency (102/1 and 34/1) were treated identically for the first four trials blocks, and it was surprising that they were dissimilar in performance. Group 102/3, with 33% relative frequency, showed somewhat larger errors and slower improvement than Group 102/1 (100% relative frequency) over the course of trials, which is consistent with the earlier literature showing that fewer KR administrations slows improvements in acquisition. These same general trends were found for VE as well. In the acquisition phase, decreased relative frequency resulted in larger errors and generally slower improvement with practice.

Results--immediate transfer phase. The immediate (10 min) no-KR transfer tests are shown for |CE| in the left portion of Figure 2, where the scores are presented in 5-trial blocks. Group 34/1 showed greater errors than either 102/1 and 102/3, which were very similar, but these group effects, $F(2,45)= 1.6$, and the groups by blocks interaction, $F(8,180)<1$, were not significant, $ps>.05$. In general, then, increased practice produced more effective performance in immediate transfer, but reducing the relative frequency from 100% to 33% did not affect performance.

For response consistency (VE), there was again a tendency for Group 34/1 to have slightly more inconsistency than Groups 102/1 and 102/3 which were about equivalent. This conditions effect was significant, with $F(2,45)=3.4$, $p<.05$. That the two 102-trial conditions were similar suggested that reduced relative frequency did not degrade performance in immediate transfer.

Results--delayed transfer test. The delayed (2-day) no-KR transfer performances for |CE| are also shown in Figure 2 (right). After two days, there was additional retention loss (as compared to the 10-min transfer test), with an approximately 30% greater error after two days than after 10 min.
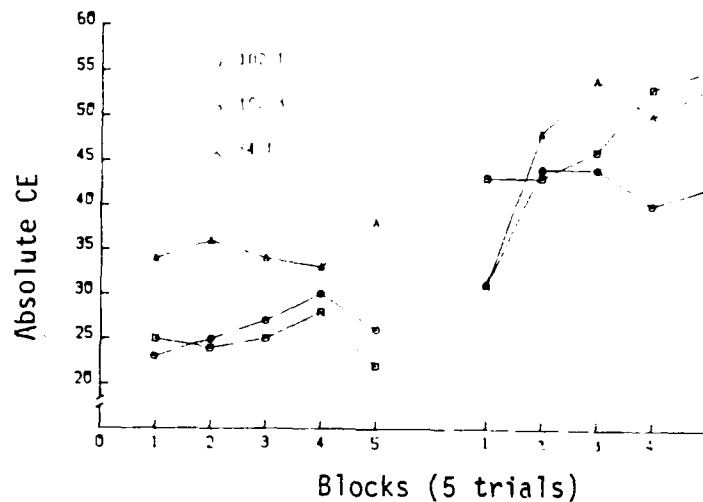
Figure 2. Mean absolute constant error (|CE|) for the immediate (1-min, left) and delayed (2-days, right) no-KR transfer tests in Experiment 1A; groups are labeled as in Figure 1.

The ordering of the groups appeared to shift across blocks, with the 100% relative frequency conditions having small errors on the first block and showing considerable decrement across blocks, with the 102/3 condition being relatively stable. There was a trend for less error overall for the 102/3 condition, but this effect was not significant, $F(2,45)<1$. The conditions by blocks interaction was *not significant either*, $F(8,180)=1.4$, $p>.05$.

For VE, there was considerable reduction in variability across blocks, $F(4,180)=16.3$, $p<.05$, which was unusual given the lack of KR in the transfer test. But all groups appeared to perform with similar trends, and the conditions effect was not significant, with $F(2,45)=1.1$, $p>.05$. Relative frequency variations in acquisition appeared not to affect consistency of delayed transfer performances.

Discussion. These results provide only weak support for the hypothesis that lower relative frequencies in acquisition would produce more effective learning, as evidenced on the two transfer tests. There was a tendency for the 33% condition (102/3) to show more effective delayed transfer performance, but this effect was not significant, and it was not present in the immediate transfer. We find this trend encouraging, however, and will examine it again in further experiments using more complex tasks.

These results are somewhat like the work on animal learning, where smaller relative frequencies of reinforcement degrade performance in training, but enhance responding on an extinction test. (This similarity is not present in our Experiment 1C, however, where immediate feedback degraded retention performance.) However, viewed from the perspective of human motor learning, these results tend to contradict expectations from most theories of motor acquisition, in that providing less KR (i.e., a 33% relative frequency) tended to improve, or at least did not significantly reduce,

retention performance relative to a 100% relative frequency condition. It raises interesting questions about how the trials without feedback can contribute to the learning process, as the common viewpoint is that such "blank" trials will not have any effects. Alternatively, it is possible that the processing of feedback information when it is occasionally provided (in Group 102/3) is somehow more effective than in Group 102/1 where it is always given. These questions will be the focus of some of our future work.

Experiments 2A and 3A

Two additional experiments are currently underway examining some of these relative frequency effects again in more complex tasks. These are being conducted as part of a Ph.D. thesis in our laboratory by Carolee Winstein. Her first experiment contrasts four different relative frequency conditions in acquisition, and uses a variety of relative frequency transfer conditions. She also studies transfer to slightly different tasks as tests of generalization, and asks about the subjects' capabilities to detect their own performance errors. A much more complex task is used, described in the last section here, which involves learning a lever movement with a particular spatiotemporal pattern. At present, one experiment is completed and awaiting a full data analysis, and another is in data collection stages.

Summary KR Effects (Paradigm B)

Background

A second major paradigm for our project involves the summary KR paradigm, where KR is provided only after a set of trials (e.g., 10) has been completed. These summaries give information about the performance on each of the trials summarized, providing exactly the information that would be received if KR was given after each trial, but the subject must wait for the information until the set of trials has been completed. Lavery (1962) studied these effects with very simple motor tasks, comparing Immediate KR, Summary KR, or Both, in an acquisition phase, and in no-KR transfer tests on subsequent days. In acquisiton, performance of the Summary conditions was markedly poorer than either the Immediate or Both conditions which were not appreciably different, with the rate of improvement of the summary condition being slowed. But in the no-KR transfer tests on subsequent days, the Summary groups tended to maintain performance, whereas the Immediate and Both conditions suffered large retention losses, so that the most effective condition in transfer was the group that had learned with the summary KR conditions. This is another situation in which a variation of KR which degrades performance in acquisition improves performance in a retention test without KR.

Because summary KR is a variable which (a) can be manipulated in real training situations run by the Army, and (b) appears to influence retention of these simple skills, we proposed to examine these effects somewhat further with more complex tasks, with more practice, and with some improvements in experimental design.

Experiment 1B

The first experiment in this area was designed to repeat and extend Lavery's (1962) findings, but using somewhat more practice, and a slightly more complex movement task. We also manipulated the number of trials summarized to search for an optimal summary length.

Procedures. Undergraduates at UCLA, 36 males and 36 females, participated for class credit. The task was again our ballistic timing task, exactly as described previously for the relative-frequency paradigm, with a movement time goal of 1000 msec. All subjects received 140 trials of practice in an acquisition phase, and then received no-KR transfer tests of 30 trials each after 10 min or two days.

Four treatment conditions were formed based on the number of trials (1, 5, 10, or 15) used in the summary KR during the acquisition phase--labeled Group 1, 5, 10, and 15, respectively. After each trial, the experimenter recorded the subject's movement time as a data point on a graph of movement time against trials. The data points were written with a grease pencil on transparent plastic which covered the graph paper, allowing them to be erased easily with a soft cloth. Regardless of the number of trials summarized, the plot always began at the left end of the abscissa with Trial 1. After the appropriate number of trials had been completed, the experimenter showed the graph to the subject for 10 sec, allowing KR about all of the trials in the set summarized. A 1-trial summary KR was essentially the "usual" immediate-KR condition, but in which the KR was given via the graph for a single trial instead of orally. The intertrial interval was 10 sec for trials *without KR, and was increased to 20 sec for the presentation of the summary information, with the* graph being presented approximately 5 sec after the last trial in the set to be summarized.

Error measures. As before, the movement times for each of the trials were grouped into blocks of trials, and various within-subject statistics (E,VE, CE, and |CE|) were computed. Only the major components of the proficiency VE and |CE| are reported here.

Results--acquisition phase. Figure 3 contains the mean |CE| scores, which measure the absolute amount of directional bias, over 10-trial blocks for the various treatment conditions in acquisition. In general, the condition with the smallest error, and the most rapid gains in proficiency, was Group 1, which demonstrated approximately asymptotic performance by Block 3. The remaining groups showed somewhat larger errors and slower rates of improvement throughout practice, with the errors being larger as the number of trials summarized was increased. Across blocks, the group means were 22.3, 31.8, 48.4 and 52.1 msec for Groups 1 through 15, respectively. This conditions effect was significant, $F(3,68)=10.8$, $p<.05$, as was the conditions by blocks interaction, $F(51,1156)=3.3$, $p<.05$. Gender effects were not significant, $F(1,64)=1.9$, $p>.05$, nor were interactions with gender. Clearly, increasing the number of trials summarized decreased directional bias and slowed the rate of improvement in acquisiton.
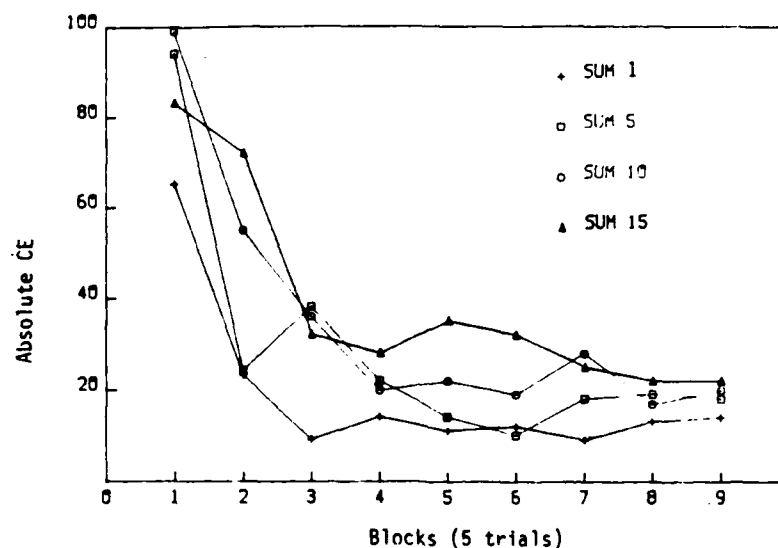
Figure 3. Mean absolute constant error (|CE|) for 5-trial blocks of trials in acquisiton as a function of the number of trials included in the KR summaries (Experiment 1B).

The effects were considerably different in the analysis of VE, the measure of response consistency. As the number of trials summarized increased, there was a tendency for the subjects to become somewhat more consistent in their responding, with the overall group means being 30.8, 26.6, 28.5, and 26.0 msec for Groups 1 through 15, respectively. This conditions effect was somewhat unstable, however, and was not significant, $F(3,68)=1.5$, $p>.05$, nor was the conditions by blocks interaction, $F(51,1156)=1.2$, $p>.05$. Gender effects were again not significant, $F(1,64)=.01$, $p>.05$, nor were interactions with gender. There was some tendency for subjects to perform somewhat least consistently in the 1-trial summary condition, perhaps because KR on each trial provided a basis for making trial-by-trial changes in behavior. When KR was delayed by 5 or more trials, such trial-by-trial alteration were perhaps not possible.

Results--immediate transfer tests. The mean |CE| scores for the immediate no-KR transfer tests, over 5-trial blocks, are shown in Figure 4 (left). All groups had regressed slightly in performance from the acquisition phase. The ordering of treatment groups here was generally unsystematic across blocks, and the conditions effects were not significant, $F(3,64),<1$. The conditions by groups interaction was significant, however, $F(12,256)=2.3$, $p<.05$, perhaps because Groups 1 and 5 tended to increase their errors across blocks, whereas Groups 10 and 15 tended to decrease them. Gender effects were not significant, $F(1,64)<1$, nor were any interactions with gender. Thus, while there was some tendency for the groups with the largest summaries to be more stable across blocks, there were no significant differences in proficiency.
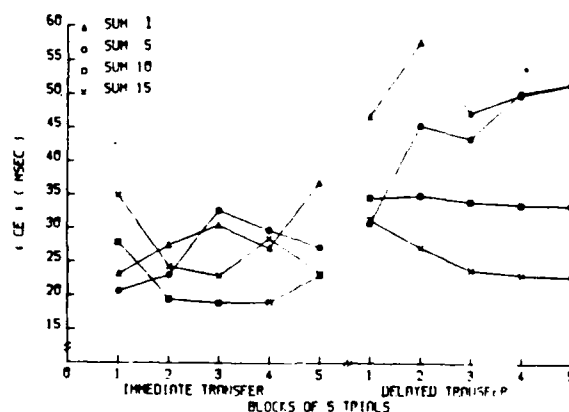
Figure 4. Mean absolute constant error (|CE|) for immediate (10-min, left) and delayed (2-days, right) no-KR transfer tests in Experiment 1B as a function of the number of trials included in the KR summaries.

Using VE as a dependent measure, there was a tendency for Group 1 to have a slightly larger errors than the other groups, but the differences were very small (ranging from 22.9 to 20.1 msec) and were not significant, $F(3,64)<1$. The conditions by blocks interaction, and the effect of gender were also not significant, $Fs<1$. Thus, for both VE and |CE|, the effects of the summary-KR manipulations were nearly totally absent across groups in immediate transfer.

Results--delayed transfer test. In the delayed no-KR transfer test, however, a very different pattern emerged, as seen in Figure 4 (right). The |CE|s were systematically larger as the number of trials summarized decreased, ranging from 25.6 msec for Group 1 to almost twice that (50.6 msec) for Group 15, with the other groups showing a systematic ordering between them. This conditions effect was significant, $F(3,64)=2.8$, $p<.05$. Furthermore, there was a tendency for Group 1 and (especially) Group 5 to increase their errors across blocks, with Group 10 remaining stable, and Group 15 decreasing its error across blocks. This conditions by blocks interaction effect was not quite significant, $F(12,256)=1.6$, $p=.08$. The gender effects were again not significant, $F(1,64)<1$, nor were there any interactions with gender. Thus, whereas increasing the number of trials summarized degraded performance in an acquisition phase, it systematically enhanced performance in a 2-day retention test without KR.

For VEs, there was a weak tendency for the 1-trial summary condition to show somewhat larger VEs than the longer summary conditions (24, 21, 21, and 21 msec, respectively), but this effect was not significant, $F(3,64)<1$. This effect was large for the males (9.5 msec) and absent for the females, but the gender effect [$F(1,64)=2.8$, $p=.1$] and the gender by conditions interaction [$F(3,64)=1.7$, $p>.05$] were not significant. Unlike the |CE|s which showed large effects of summary conditions, the VEs were relatively unaffected.

Discussion. The summary-KR manipulations degraded performance markedly in the acquisition phase, increasing errors and retarding the rate of gains in proficiency. In our no-KR

transfer tests, no effect of these variables on either |CE| or VE were found in the immediate (10-min) transfer test. But in the delayed test (2 days), increasing the number of trials summarized resulted in systematically smaller |CE|s, with unchanged VEs. Thus, surprisingly, summary KR appears to be a variable which degrades performance in acquisition (relative to immediate KR), but which improves proficiency in a delayed no-KR transfer test. These differences were very large, and were relatively stable across blocks.

Theoretically, these effects could be related to a number of processes. First, the summary-KR manipulations may have affected the extent to which subjects learned the task during the acquisition phase, perhaps by requiring additional information processing activities that were not required by the group with relatively immediate feedback. If so, one would have perhaps expected to find that these effects were manifested on the immediate retention test, when the temporary effects of the KR had been supposedly removed. The effects did not appear, however, until a delayed (2-day) retention test, suggesting the alternative possibility that the summary-KR conditions could have in some way rendered the subjects less susceptible to forgetting, with the level of learning at the beginning of the retention interval being similar across groups. At face value, the effects seen in Figure 4 seem to favor this explanation, because there was (a) no difference among groups on an immediate retention test, (b) large differences on the delayed test, (c) little nor no retention loss for Group 15 from acquisition to delayed transfer, and (d) systematically larger losses for the groups with larger numbers of trials summarized.

At the same time, the question arises as to what may have been learned, or what has become more resistant to forgetting, as a result of these summary KR procedures. One possibility is that the basis of the effect is a more effective movement control (i.e., a "stronger" motor program) created as a result of the conditions in acquisition. This is possible, but if so one would have expected to find the effects manifested in variable error (VE)-- the statistic usually thought to be most sensitive to the development of motor programming capabilities (e.g., Schmidt, 1982)--and not manifested by |CE| which is a measure of directional bias or "drift" from the goal movement time. Another hypopthesis, though, is that the summary-KR procedures generated a more effective error-detection capability, perhaps by forcing the learner to focus on his/her own feedback after each trial, and leading the subjects to practice the additional task of estimating their own errors. If so, and if this capability for estimation increased with practice, then such added capabilities would be effective in maintaining performance in no-KR transfer tests where no external evaluations are present. Also, such an explanation could explain why Group 15 improved their performance during the delayed transfer test, whereas Groups 1 and 5 showed increased error, as if they were drifting off target and did not know it. These hypotheses are under test in some of our experiments planned for the coming year, and in the experiments to be reported in Paradigm C (Subjective Estimation) below.

These somewhat counter-intuitive summary-KR effects will require much more study before they can be adopted as stable principles of motor learning, of course, but even now they provide

implications for Army training. If the criterion of success is a delayed transfer test with degraded feedback (often the case in riflery, for example), then summary KR procedures might have strong relevance for Army training settings of this type. Also, unlike many laboratory manipulations, summary KR as a method for providing feedback seems easily implemented in many actual training situations used by the Army. Of course, we do not yet know how long such effects may persist, whether or not they generalize to other, more complex movement tasks with more relevance to the Army, what an optimum summary length might be, and whether or not such an optimum might interact with the nature of the task in some way. Many of these questions are addressed by experiments already underway in our second year.

Experiment 2B

At present, we have begun another experiment in this series, in which these effects of summary KR procedures are sought in a somewhat more complex coincident-timing task (see the Paradigm Development Section here), using exactly the same experimental design as in Experiment 1B. At this writing, the data are nearly collected.

Subjective-Estimation Effects (Paradigm C)

Background

Some years ago, a number of us (e.g., Schmidt & White, 1972) used a paradigm in which the subjects were asked to guess their own errors after a trial, before KR was given; the accuracy of this subjective estimation was used to determine the strength of error-detection capabilities which were thought to be developed during practice (see Schmidt, 1982, Chapter 14). However, Hogan and Yanowitz (1978) examined the question in reverse: they asked whether such subjective-estimation activities could in some way increase movement proficiency in the task. Asking (or not asking) subjects to estimate during acquistion had little effect on performance. But in an immediate no-KR transfer test, subjects who were asked to estimate maintained performance across trials almost perfectly, whereas those subjects not asked to estimate increased their errors markedly. These data suggested that subjective estimation contributed to the learning of movement proficiency in some way.

These results seemed to be strongly tied to our earlier data concerning various KR manipulations. One hypothesis is that, in the conditions with low relative frequency or long summaries where KR is not presented for a string of trials, subjects use a strategy in which they attempt to estimate their own KR. One possible procedure would be to attend to their own response-produced feedback (audition, kinesthesis)--present even though KR may be withheld--as a basis for this guess. If so, then one would expect subjects to learn a relationship between response-produced feedback and movement time--what we and others have termed an error-detection mechanism (Schmidt & White, 1972; Schmidt & Wrisberg, 1973; Schmidt, 1975). Such a capability to detect their own errors would then be available on no-KR transfer trials, and would allow subjects to maintain performance, or perhaps even to continue to learn, using their

subjective errors as a basis. Thus, the Hogan-Yanowitz paradigm allows the study of motor learning as a function of these enforced estimation procedures in acquisition.

The Hogan-Yanowitz study has strong potential generalization to Army training settings, mainly because (a) it manipulated variables easily implemented in Army training and (b) it focused on retention of learned capabilities as a function of these variables. The study had a potential problem, though. The transfer test did not equate the independent variable (estimate vs. no-estimate) across conditions as is usually done in analyses of learning vs. performance effects; rather, subjects who estimated in acquisition continued to do so in transfer, and vice versa. This experiment needed to be conducted again, using equated estimation conditions in transfer. Also, there were relatively few practice trials in acquisition (50), and somewhat more practice would have been useful. And, an experiment in which both estimation and no-estimation conditions were used in transfer, and with both immediate and delayed transfer tests, would add considerably to the generality of these effects. These rationale served as the basis for the third paradigm in our project, and four experiments which focus on such questions are described below.

Experiment 1C

Our initial experiments, conducted with collaboration with Stephan Swinnen from Belgium, attempted to generalize and extend the Hogan-Yanowitz study in various ways. We used more practice, larger sample sizes, and various retention tests for which the independent variable in acquisition was controlled.

Procedures. Undergraduates at UCLA, 32 males and 32 females, participated for class credit in a course in Kinesiology. The apparatus and task were exactly as described before, with the goal movement time being 1000 msec. Subjects received 120 trials of practice with KR in an acquisition phase, and 30-trial no-KR transfer tests after 10 min and 2 days.

The experiment involved a simple 2 x 2 design, with estimation or no-estimation conditions in acquisition crossed with estimation or no-estimation conditions in the transfer tests, forming four groups of 15 subjects each (7 males and 8 females in each). In the no-estimation conditions, subjects produced a movement, waited for 9 sec until KR was provided, and then prepared for the next trial 6 sec later; the intertrial interval was 15 sec. For the estimation conditions, the subjects were asked to guess their movement time 4 sec after making the movement, then KR about the actual movement time was presented 5 sec later, so that KR delay and the intertrial interval were equated across conditions.

Dependent measures. As before, various within-subject measures were computed for 5-trial blocks (E, VE, CE, and |CE|). VE as a measure of consistency and |CE| as a measure of bias are reported here.

Results--acquisiton phase. In acquisiton, average |CE| dropped considerably across practice, especially through Block 6 (i.e., Trial 30), and continued to decrease slightly thereafter such that the average bias was approximately 20 msec by the end of practice. However, there were no obvious

differences between the groups asked to estimate and the groups not asked to estimate, and this effect was not significant, $F(1,62)=2.1$, $p>.05$. A similar pattern was evidenced using VE as a dependent measure, with VEs dropping from about 200 msec in the first block to around 18 msec at the end of practice, but with the majority of the improvement occurring prior to Block 10. There was a slight (5 msec) tendency for the estimation group to have slightly greater VE than the no-estimation group, as if the requirement to estimate were disrupting performance in some minor way, but this effect was not significant, $F(1,62)<1$. Generally, there was no evidence that enforced estimation in the acquisition conditions influenced performance for either dependent measure.

Results--immediate transfer phase. The four conditions formed by crossing the two conditions in acquisition and the two conditions in transfer are shown in Figure 5 (left), using |CE| as the dependent measure. Subjects asked to guess in acquisition (open symbols) performed with slightly greater error than subjects not asked to guess in acquisition (filled symbols). This seemed to be evident both for subjects asked to estimate in the transfer test (squares) and the subjects not asked to estimate in the transfer test (circles). However, effects of estimating in acquisition [$F(1,60)=1.2$, $p>.05$], of estimating in the transfer test [$F(1,60)=1.4$, $p>.05$], and the acquisition conditions by transfer conditions interaction [$F(1,60)<1$], all failed significance. Therefore, requiring the subjects to estimate in the acquisition or in the transfer phase had no effect on |CE| in immediate transfer.
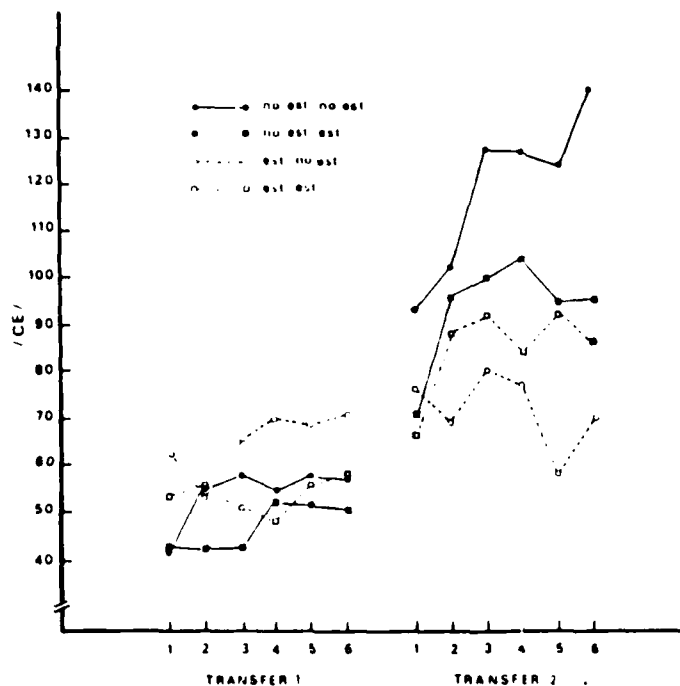

Figure 5. Mean absolute constant error (|CE|) over 5-trial blocks in the immediate (10-min, left) and delayed (2 days, right) no-KR transfer tests in Experiment 1C. (Filled symbols=no estimation in acquisition, open symbols=estimation in acquisition; circles=no estimation in transfer; squares=estimation in transfer.)

The effects for VE showed a similar trend, with a tendency for subjects who estimated in acquisition to have a slightly larger VE than subjects who did not (39 vs. 32 msec), as if the slightly (nonsignficantly) depressing effects of estimation in the acquisition phase carried over into the immediate transfer test. This effect was significant, $F(1,60)=8.6$, $p<.05$. This trend was in a general way opposite to that found by Hogan and Yanowitz (1978), who found beneficial effects of estimation in acquisition using absolute error scores in immediate transfer. Neither the effect of estimation in transfer [$F(1,60)<1$] nor the acquisition conditions by transfer conditions interaction [$F(1,60)<1$] were significant, $p<.05$. Thus, requiring subjects to estimate in acquisition increased VE slightly in immediate transfer, but the difference was not particularly large as compared to the effects seen in delayed transfer, presented next.

Results--delayed transfer phase. In the delayed (2-day) transfer test, shown for |CE| in the right portion of Figure 5, the differences among groups were much larger than in immediate transfer, caused mainly by substantial retention losses for all groups across the 2-day interval. Overall, subjects asked to estimate in the acquisition phase (open symbols) performed with less average bias than subjects not asked to estimate in acquisiton (filled symbols), which was significant, $F(1,60)=4.2$, $p<.05$. This difference was especially large for the subjects who did not estimate in transfer (open vs. filled circles), and was smaller for the subjects who estimated in transfer (open vs. filled squares), but the acquisition conditions by transfer conditions interaction was not significant, $F(1,60)=2.1$, $p>.05$. The effect of estimation in the transfer test was not significant, $F(1,60)<1$. Furthermore, for delayed transfer, the effect of estimation in acquisition on delayed transfer performance was much larger for males than for females; the no-estimation acquisition condition had nearly twice the error as the estimation acquisition condition for the males, but there was essentially no effect for the females. In a separate ANOVA including gender as a factor, females tended to have less error than males, but this effect just failed significance, $F(1,56)=3.0$, $p=.08$. But the interaction between conditions in acquisition and gender was signficant, $F(1,56)=4.8$, $p<.05$. No other interactions with gender were reliable, $Fs(1,56)<1$.

For VE as a dependent measure, these effects of conditions in acquisition were nearly completely absent. The effects of acquisition conditions, transfer conditions, and their interaction were not reliable, $Fs(1,60)<1$. As well, in a separate analysis, gender [$F(1,56)<1$] and all interactions with gender [$Fs(1,56)<1.1$, $ps>.05$] were nonsignficant. Apparently, none of the effects of estimation was manifested in terms of response consistency, VE.

Overall, for delayed transfer, the effect of estimation in acquisition was to decrease the bias (i.e., |CE|), with the effect being far larger for the groups not asked to estimate in transfer. Furthermore, this effect was mainly attributable to the male subjects. Only minimal, somewhat inconsistent, effects were found for response consistency (VE).

Discussion. This experiment provided some evidence that requiring learners to estimate their

errors in acquisition leads to increased peformance on delayed retention tests without KR. But a number of features of these results cause us to be somewhat cautious in accepting them at face value. First, the effect was only present for male subjects--in fact was totally absent for female subjects--leaving serious doubts about generalizability. Second, the effects were only present for our measures of response bias ($|CE|$), and were generally absent for measures of response consistency (VE). This is disturbing, in that VE is usually taken as the most "important" measure of motor proficiency, representing the subjects' fundamental inability to reproduce skilled behavior on successive trials; and requiring considerable practice to be reduced. On the other hand, response bias measures are thought to be easily created and reduced with very little practice; it is as if the subjects' "calibrations" were slightly off, a trial or two being needed to return them to optimum settings. Such effects can be important, of course, as they affect aspects of responding such as "aim" in riflery. But they diminish in importance for training if they can be easily eliminated with slightly more KR practice.

The effects were present mainly on a delayed retention test, with no major effects being shown after a 10-min delay. Such effects were different than in Hogan and Yanowitz' (1978) experiment, where beneficial effects of guessing in acquisition were found (using absolute error) on an immediate no-KR transfer test (no delayed test was given). However, Hogan and Yanowitz did not provide a 10-min retention interval as we did; they required subjects in the no-estimation conditon to recite a trigram as a control for attentional requirements; but perhaps most important, they did not transfer the subjects to equivalent levels of the estimation variable in transfer, but allowed the estimation condition to continue estimating, continuing to prevent it for the no-estimation condition. Any of these differences could be the cause of the disagreement here.

Our results suggest that enforced estimation in the acquisition phase probably did not act to enhance learning of the task, because such improved capabilities for responding would have been evidenced immediately. However, forced estimation in acquisition could have acted to prevent retention losses in some way, making the subjects less susceptible to extra-experimental influences that would cause them to bias their responses in transfer. In addition, the estimation conditions in acquisition seemed to allow the subjects to maintain performance more effectively in delayed transfer; compare the open symbols (estimation in acquisition) in Figure 5 (right) with the closed symbols (no estimation in acquisition). Notice that all groups (exception: solid circles) began at about the same level of performance, but the groups with estimation in acquisiton regressed considerably less than those without estimation in acquisition. Such an effect suggests that the estimation contributed to the development of the capability for subjects to detect their own errors, and that subjects with a weaker capability drifted from the target, apparently unaware.

These effects were sufficiently interesting at this point to lead us to investigate them further. We felt that a first step should be a replication and extension of these effects, which was the rationale for the second study in this series.

Experiment 2C

In this study, we attempted to replicate the effects shown in Experiment 1C, but with a task having a shorter movement time goal. Many have argued (e.g., Schmidt, 1975, 1982) that mechanisms to produce an action (motor programs) are more easily separated from mechanisms that evaluate its correctness if the movement is rapid as opposed to slow. If so, then if the basis of the effect in Experiment 1C was in terms of more effective response evaluation mechanisms, this effect should be enhanced with the more rapid task used here.

Procedures. Undergraduates at UCLA, 32 males and 32 females, participated as a requirement for a course in Psychology. None had served in Experiment 1C. The apparatus and task were exactly as in the previous study, except that the movement time goal was 550 msec instead of 1000 msec, only 80 trials were used in the acquisition phase, and two transfer tests of 25 trials were given (rather than 30). Again, we had two conditions (estimation and no-estimation) in acquisition crossed with estimation and no-estimation conditions in transfer. As before, no-KR transfer tests were conducted after 10 min and 2 days. All other aspects of the experiment, such as instructions, timing of events within a trial, dependent measures, etc., were identical with those of Experiment 1c.

Results--acquisiton phase. As in the previous study, there was considerable decrease in |CE| with practice, with average bias decreasing from approximately 150 msec in the first block to approximately 15-20 msec in later blocks, and the majority of improvement occurring through Block 10 (i.e., Trial 50). There was no obvious difference between the estimation and no-estimation conditions across practice, and the conditions effect [$F(1,62)<1$] and the conditions by blocks interaction [$F(15,930)<1$] both failed significance, $p>.05$.

Measures of response consistency (VE) decreased considerably across practice as well, decreasing from 85 msec in the first block to approximately 25 msec in the final block, with the majority of decrease occurring before Block 6 (i.e., Trial 30). There was a small tendency for the estimation condition to have slightly less VE than the no-estimation condition (33 msec vs. 30 msec--opposite to the trend found in Experiment 1C), but this effect just failed significance, $F(1,62)=3.2$, $p=.08$. The conditions by blocks interaction was not reliable either, $F(15,930)<1$.

Overall, in acquisition, the estimation and no-estimation conditions appeared to be similar in performance for both |CE| and VE. This pattern was similar to that found in Experiment 1C.

Results--immediate transfer phase. Figure 6 has the average |CE| scores for the four groups, showing the immediate (10-min) no-KR transfer performances at the left. The groups which estimated in acquisition had slightly larger errors than the groups which did not (31.6 vs. 30.3 msec, averaged across blocks), but this acquisition condition effect failed significance, $F(1,60)<1$. On the other hand, the groups which estimated in the transfer test showed considerably less error than the groups that did not estimate in the transfer test (26.6 vs. 35.4 msec), which was significant,
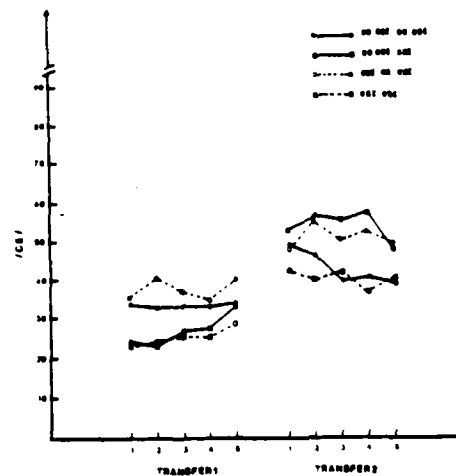
Figure 6. Mean absolute constant error (|CE|) for 5-trial blocks in the immediate (10-min, left) and delayed (2 days, right) no-KR transfer tests for Experiment 2C. (Group labels are as in Figure 5.) some way.

$F(1,60)=4.1$, $p<.05$, suggesting that forcing estimation in transfer was beneficial to performance in The interaction between acquisition and transfer estimation conditions was not significant, $F(1,60)<1$.

For VE in the immediate transfer phase, the groups which estimated in acquisition showed slightly smaller VE than the groups which did not estimate in acquisition (20.3 vs. 24.3 msec), and this effect was signficant, with $F(1,60)=6.3$, $p<.05$. But the estimation conditions in the transfer test, and the acquisiton estimation condition by transfer estimation condition interaction, were not significant, with $Fs(1,60)<1$. Also, in a separate analysis including gender as a factor, there was a significant three-way interaction between gender, acquisition condition, and transfer condition; essentially, the beneficial effect of estimation on immediate transfer performance was larger for the females (19.7 vs. 25.7, or 6 msec) than for the males (20.9 vs. 22.9, or 2 msec), $F(1,56)=4.7$, $p<.05$. Thus requiring estimation in acquisition produced more consistent responses in the immediate transfer test, especially so for the female subjects, and regardless of the estimation conditions required in transfer. These effects were opposite to those found in Experiment 1C, where estimation produced increased VE in immediate transfer.

Results--delayed transfer test. For the average |CE| scores in the delayed (2-day) no-KR transfer test shown in the right portion of Figure 6, there was some tendency for the groups which estimated in acquisition (open symbols) to have smaller bias than the groups which did not (filled symbols), but this difference (45.4 vs. 48.5 msec) was not significant, $F(1,60)<1$. Also, neither

the effect of the estimation conditions in transfer [$F(1,60)<1$] nor the interaction between the estimation conditions in acquisiton by estimation conditions in transfer [$F(1,60)=3.3$, $p=.07$] were reliable. However, in a separate analysis including gender as a factor, there was a significant interaction between acquisition conditions and gender, $F(1,56)=5.9$, $p<.05$. This effect, where estimation in acquisition was detrimental to females (53.5 vs. 42.3 msec) but was beneficial to males (37.2 vs. 54.7 msec), was essentially _opposite_ to that found in Experiment 1C. The analysis of |CE| shows no evidence for beneficial effects of estimation in acquisition on delayed transfer performance.

For response consistency (VE), groups which estimated in acquisition had slightly more consistent performances in delayed transfer than did groups which did not estimate in acquisition (23.4 vs. 25.9 msec), but this effect was not significant, $F(1,60)=2.5$, $p>.05$. The effect of the transfer test estimation conditions was not significant, nor was the acquisition estimation condition by transfer estimation condition interaction, $Fs(1,60)<1$. In the separate analysis involving gender, this factor did not produce any significant main effects or interactions with the other factors, $Fs(1,56)<1$. For VE, there was no evidence that estimation conditions in acquisition affected delayed transfer performance.

Discussion. This experiment generally failed to replicate the effects found earlier. In particular, estimation in acquisition reduced VE scores in the immediate transfer test, essentially opposite to the effect found in Experiment 1C. But more importantly, in the delayed transfer test, there was only small beneficial (nonsignificant) effects on response bias measures |CE| in favor of the group which had estimated in acquisition. Further, the gender effects in Experiment 1C, with the males showing the effects of acquisition more than females, was reversed in Experiment 2C. Clearly, the effects of estimation seen in Experiment 1C have not been replicated in any convincing way.

The experiments were different, however, in minor ways. First, the task was more rapid in Experiment 2C. We had expected to find the estimation effects to be larger with the more rapid task, not smaller as was found. Also, the subjects were from a different population. In Experiment 1C, the subjects were Kinesiology Department undergraduates, who were participating to gain extra class credit. In Experiment 2C, on the other hand, subjects were from the Psychology Department Subject Pool, in which students are required to participate as a part of a course in Psychology. Subjects in Experiment 2C (Psychology undergraduates) seemed less "motivated" (in the experimenter's judgement) than subjects in Experiment 1C (Kinesiology undegraduates), but there is no objective evidence to support this claim. Whether or not these factors are responsible for the differences in outcome between these two experiments is unclear and we are in a particularly weak position to say very much conclusive about the role of subjective estimation in acquisition on motor learning or retention.

These two experiments examined the role of estimation by attempts to _enhance_ subjective estimation over and above a no-estimation "baseline" condition. In the following experiments, we

attempted to block subjective estimation in one condition, by providing KR for the subjects instantly after a trial. If KR is provided immediately, we reasoned that subjects would not engage in subjective estimation, as the product of that estimation had already been given to them in the form of KR. If an immediate-KR condition has subjective estimation processes blocked, then perhaps this would degrade the progress of learning, and would be manifested on various transfer tests without KR. This was the rationale for Experiment 3C.

Experiment 3C

Procedures. University undergraduates, 39 males and 39 females, participated in exchange for extra credit towards a course in Kinesiology. The apparatus and task were exactly as in Experiment 1C, with a 1000-msec movement time goal. Subjects received 90 trials in an acquisition session with KR, and then received 30-trial no-KR transfer tests after 6 min and 2 days. Dependent measures were VE and |CE|, as before.

The experiment involved three treatment groups (13 males and 13 females in each), all with a 13-sec intertrial interval. For two conditions, KR was given 8 sec after completion of the movement, with a post-KR delay interval of 5 sec. One group (Estimation) was required to guess their error in milliseconds prior to the presentation of KR; the other (No-Estimation) simply rested in the KR-delay interval. A third group (Immediate) received a 0-sec KR delay. A digital millisecond timer was positioned just above the movement's endpoint, and subjects in this condition could see the movement time from the timer immediately as it stopped when the switch was tripped. Subjects in the other two conditions viewed KR presented on the timer as well, but the experimenter delayed its viewing by placing a shield over it until the 8-sec KR-delay interval had elapsed. (The numerals were "blurred" when the timer was running, which probably prevented meaningful timing information from being used during the movement.)

Results--acquisition phase. The performances during acquisition resembled those from other experiments here, with |CE| decreasing from about 205 msec in the first block to about 40 msec in the 18th block. All groups had about the same performance over the last half of the acquisition phase, and the conditions effect was not significant, $F(2,75)<1$. However, there were some differential trends across trials, with the Immediate group appearing to decrease their errors most rapidly in the first few blocks, and the Estimation group least rapidly. This was probably the cause of the significant conditions by blocks interaction, $F(34,1275)= 1.6$, $p<.05$. In general, while the overall performance levels were about the same across groups, the Immediate group appeared to reach this level somewhat sooner, and the Estimation group reached it later.

For VE, trends similar to those just described for |CE| were found. There was no effect of conditions averaged across blocks, $F(2,75)<1$, but the conditions by blocks interaction was significant, $F(34,1275)=1.8$, $p<.05$. Again, the Immediate group appeared to reach the asymptote somewhat more quickly than the No-Estimation condition, with the Estimation condition being slowest.

Results--immediate transfer. The averaged |CE| scores for the immediate transfer test are shown in the left portion of Figure 7. All groups regressed in performance considerably relative to the end of acquisition. The performances across the five blocks were somewhat irregular, but there was a tendency for the No-Estimation and Estimation conditions to perform approximately the same, but with the Immediate group showing somewhat larger error, together with more loss in error across blocks. The effect of acquisition conditions was not significant, however, $F(2,75)<1.0$. The conditions by blocks interaction however, fell just short of significance, $F(10,375)=1.8$, $p=.06$, tending to support the differential trends over blocks seen in Figure 7. Thus, there was some tendency for the Immediate condition to produce the least effective performance in immediate transfer.

These effects appeared to be generally present for response consistency (VE), with a somewhat larger overall error for the Immediate group relative to the other two which were about equivalent. However, the conditions main effect was not significant, $F(2,75)<1$. There was also a trend for the Immediate group to lose proficiency more rapidly than the other groups, but the conditions by blocks interaction was not significant, $F(10,375)=1.5$, $p>.05$. Thus, for VE, there was little evidence of differential performance as a function of conditions in acquisition.

Overall, in immediate transfer, there was the suggestion using both VE and |CE| that the Immediate group lost proficiency over blocks more quickly than the other two groups. But initial performance for the Immediate group was usually equivalent to that of the other groups, and the lack of significance for the condition effects tended to weaken this view of the results considerably. It is probably best to say that performances were essentially equivalent in immediate transfer.
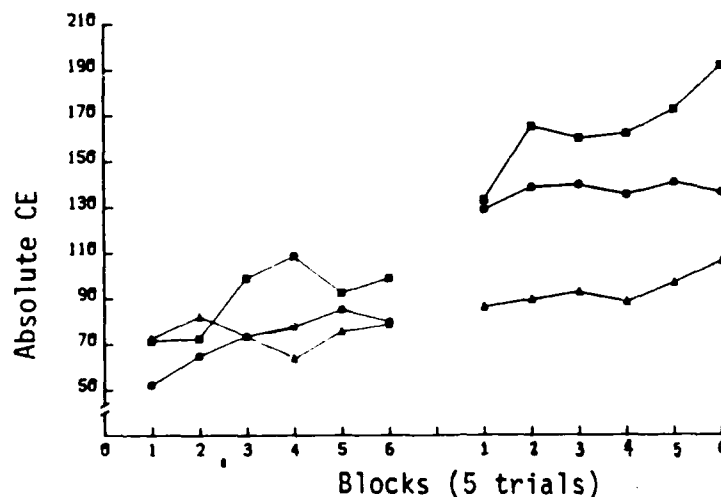

Figure 7. Mean absolute constant error (|CE|) over 5-trial blocks for immediate (6-min, left) and delayed (2 days, right) transfer for Experiment 3C. (Triangles=Estimation, circles=No-Estimation, and squares=Immediate.)

Results--delayed transfer. The mean |CE| scores for the three groups on delayed transfer are shown in the right portion of Figure 7. Again, there was further retention loss here as compared to the immediate retention test, but the losses appeared to be considerably greater for the Immediate group, with the estimation condition showing the least loss. The conditions effect here was significant, $F_{(2,75)}=3.8$, $p<.05$. Also, the groups appeared to lose proficiency across blocks at about the same rate, and the conditions by blocks interaction was not significant, $F_{(10,375)}=1.2$, $p>.05$. Response bias measures, therefore, tend to show that the Immediate condition was the least effective in terms of the delayed no-KR retention test.

For response consistency (VE), the conditions were ordered again as they were for |CE|, with the Immediate condition showing most variability (167 msec), the No-Estimation condition an intermediate value (142 msec), and the Estimation condition the lowest (103 msec). This effect was significant, with $F_{(2,75)}=3.8$, $p<.05$. The conditions by blocks interaction was not significant, $F_{(10,375)}=1.2$, $p>.05$.

Generally, in delayed transfer, the most effective performance (for both |CE| and VE) was by the Estimation condition, the least effective was by the Immediate condition, with the No-Estimation condition falling generally mid-way between them. Also, the estimation condition was generally more accurate than the no-estimation condition, as had been found in Experiment 1C (but not in Experiment 1B).

Discussion. These results support our hypothesis that preventing subjective estimation--by giving KR instantaneously, before the subject would engage in estimation spontaneously--would interfere with the acquisition of the task. This interference was not seen in the acquisition phase, because the guiding effects of KR there are so strong that the performance differences are minimized. In fact, the Immediate condition might even have provided more effective guidance in acquisition, if one views the somewhat more rapid rate of improvement there in this way. But in the no-KR transfer test, particularly when it was delayed, the Immediate condition performed very poorly, and appeared to regress in performance over trials to a greater extent than the other groups.

Our interpretation is that estimation, whether spontaneous or enforced, engages various information processing activities that act to strengthen the subjects' error detection capabilities. And, if the Immediate condition partly or completely blocked these activities, then it is not surprising that the Immediate KR condition produced the least effective long-term retention performance. This result converges with the results of Experiment 1C to suggest that these information processing activities might be important in the development of long-term retention.

Finally it is worth noting that these results are strongly counter to a number of common views of KR scheduling, where immediate feedback should be most effective. Immediate KR was most effective, if we only view learning in terms of performance in the acquisition phase. But to the extent that long-term retention is an important goal for learning--as it clearly is in Army

situations--then Immediate KR was detrimental relative to delayed KR. Notice also, that these effects of Immediate KR are opposite to the expectation for animal learning, where delayed reinforcement seriously degrades acquisition. Thus, the parallels between animal learning and human motor learning--encouraged by the results of our relative frequency paradigm perhaps (Experiment 1A)--are discouraged by the results of the present study.

As a last experiment in this series, we sought to replicate this immediate-KR effect. And, in view of the inconsistency of the effect of estimation and no-estimation, we repeated Experiment 1C, but with slightly different conditions.

Experiment 1D

Procedures. University undergraduates, 30 males and 30 females, received extra credit toward a course in Kinesiology for participating. The complex ballistic timing task was used, exactly as in the previous studies, with a goal movement time of 1000 msec. There were 120 acquisition trials, followed by two 30-trial no-KR transfer tests after 10 min and 2 days. Dependent measures were |CE| and VE computed over 5-trial blocks, as before.

The experiment involved the same three conditions as in Experiment 1C. An Immediate condition saw KR immediately from the timer, a No-Estimation condition received KR delayed by 9 sec, and an Estimation condition was asked to estimate their movement time prior to receiving KR, 9 sec after the movement. In all cases, the intertrial interval was 15 sec. Half the subjects in each group received both transfer tests, with the other half receiving only the delayed transfer test. The rationale was that our earlier studies have used subjects in both transfer tests, and we wanted to evaluate whether or not receiving the first test would influence performance on the second.

Results--acquisition phase. During acquisition, |CE| scores decreased steadily from 170 msec in the first block to about 30 msec in the final block. Overall, there was some tendency for the Estimation group to perform least accurately (53 msec), with the No-Estimation group intermediate (49 msec), and the Immediate group being most accurate (44 msec). These effects were rather unstable across blocks, however, and the conditions effect, $F_{(2,57)}<1$, and the conditions by blocks interaction, $F_{(46,1311)}<1$, were not significant. For response variability (VE), there were no obvious differences among groups, and the conditions effect, $F_{(2,57)}<1$, and the conditions by blocks interaction, $F_{(46,1311)}=1$, were not significant. Generally, there was a weak tendency for the Immediate group to be most accurate in acquisition, with the Estimation group least, but these trends were not supported statistically.

Results--immediate transfer. The mean |CE| scores for the three conditions in immediate no-KR transfer are shown in the left portion of Figure 8. The group means were very unstable across blocks, but there was some tendency for the Estimation group to have the highest errors, with the No-Estimation group least, but this effect was not significant, $F_{(2,27)}<1$. There was little evidence of group differences for |CE|.
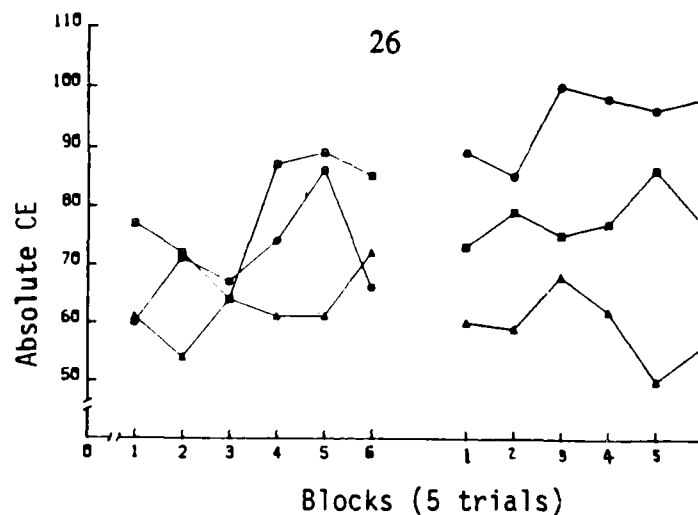
Figure 8. Mean absolute constant error (|CE|) over 5-trial blocks for immediate (10-min, left) and delayed (2-day) transfer tests in Experiment 4C. (Triangles=No-Estimation, squares=Estimation, and circles=Immediate.)

For VE, there was a tendency for the Immediate group to have least variability (33 msec), with the Estimation group being intermediate (36 msec), and the Delayed group most variability (41 msec), and this effect was significant, $F(2,27)=3.8$, $p<.05$. This effect was not particularly compelling, however, as it was generated by large differences on one block of trials, and was generally absent otherwise. There was also a signficant blocks by conditions interaction, $F(10,120)=2.4$, $p<.05$.

Overall, in immediate transfer, there was some evidence that the immediate condition was most consistent, the No-Estimation condition least consistent; no effects at all were found for |CE|. These effects have not been found in our earlier study using these methods, and the instability of the result makes us uncertain about these effects in immediate transfer.

Results--delayed transfer. The mean |CE| scores for the delayed no-KR transfer test are shown in the right portion of Figure 8. Here, as in Experiment 3C, the Immediate condition showed the greatest error (120 msec), with the Estimation group intermediate (78 msec), and the No-Estimation being most accurate (60 msec). However, this large effect was not statistically significant, $F(2,57)=2.1$, $p>.05$, probably because of unusually high within-group variabilities in several conditions. The poor delayed transfer performance for the Immediate group tends to replicate the findings of Experiment 1C, but the beneficial effects of estimation found in Experiment 1C (but not in 2C) were absent here. For VE, across all blocks the Intermediate condition was again most variable (58 msec), with the other two conditions being essentially equal (35 msec) in variability. But this effect was again not statistically reliable, $F(2,57)=1.0$, $p>.05$.

Overall, for both error measures, the Immediate condition was least effective in delayed transfer test, with the Estimation and No-Estimation conditions being essentially similar. Although these condition effects were not significant, they do tend to replicate the similar effects found in Experiment 1C. On the other hand, the comparison between the Estimation and No-Estimation

conditions was absent here, which again failed to replicate the analogous effects in Experiment 1C.

Discussion--Estimation Paradigm

For all of these experiments taken together, there was only mixed support that estimation conditions in acquisition benefit learning, as measured on various transfer tests without KR. Two experiments (1C and 3C) show this effect clearly, while attempts to replicate them (2C and 4C) have failed. These failures to replicate are both puzzling and discouraging to us, and we do not have appropriate explanations for these outcomes. But, taken at face value, the evidence does not support the hypothesis that enforced estimation (relative to no estimation) in acquisition is uniformly beneficial for training.

On the other hand, these lack of effects for estimation vs. no-estimation conditions do not necessarily mean that estimation is not a factor in learning. Rather, they could mean that these estimation procedures do contribute to learning, and that the subjects in the no-estimation control condition estimated spontaneouly. Indeed, a questionnaire conducted after an experiment (not described here) revealed that the majority of subjects in the no-estimation conditions admitted estimating even though they had no instructions to do so. If so, then perhaps error-estimation is a rather "natural" process for motivated subjects to engage in. Perhaps it is not surprising, in retrospect, that the comparison of the forced estimation condition with a so-called "no-estimation" condition produced such small effects.

This hypopthesis tends to be borne out by the results of our immediate KR paradigm. If we can accept the assumption that immediate KR actually blocks estimation processes, then the findings that the immediate KR conditions produced the poorest performance in delayed transfer tests (not significant in Experiment 4C, though) supports the hypothesis that blocking the usual (spontaneous and enforced) estimation processes in acquisiton will be detrimental to learning and retention. Our interpretation is that the subjects in the immediate condition do not attend to response-produced feedback to the extent that those in the other groups do, which interferes with the development of error-detection capabilities useful in no-KR transfer.

If so, these findings could have important implications for Army training. Care should probably be taken to ensure that learners have an opportunity to estimate their own errors. This could be done by actually asking learners to evaluate their performances in various ways. But, our results also imply that training conditions should be structured in such a way that they do not prevent learners from engaging in these self-evaluation processes. Secondary activities placed between a trial and its KR, or a too-short KR-delay or intertrial interval, might have the effect of blocking these activities in a way similar to that in our Experiments 3C and 4C here. Also, if low motivation in trainees involves a certain "detachment" from the problem or task they are practicing, it might also include a lack of effort in self evaluation, and perhaps enforced estimation would be a benefit for this type of learner.

PARADIGM DEVELOPMENT

In the previous sections, we have described experiments using very simple, laboratory motor behaviors. While much has been learned from them in this and earlier work, we are concerned that they are too simple to have generalizability to many of the behaviors to be learned in actual training settings. In this final section, we describe efforts to develop paradigms for motor learning that involve somewhat more complex movement behaviors, and yet which retain features that allow their use in laboratory settings.

Complex Spatiotemporal Patterning Task

In one of our new paradigms, modeled after early work by Armstrong (1970) and Shapiro (1977, 1978), we ask subjects to learn relatively complex spatiotemporal patterns of a lever movement with the right arm. The experimenter defines a pattern in space and time (as most real movement skills are), provided graphically on a computer terminal as shown in Figure 9. In this example, taken from dissertation work in our laboratory by Carolee Winstein, the subject is to make the movement with three reversals in direction, with the added restriction that the reversals occur at the proper time. In principle, a movement of nearly any complexity, or with any number of participating limbs, could be used. For feedback, the subject sees the template on the screen, and then the computer superimposes the just-produced movement, so that the subject can see the nature
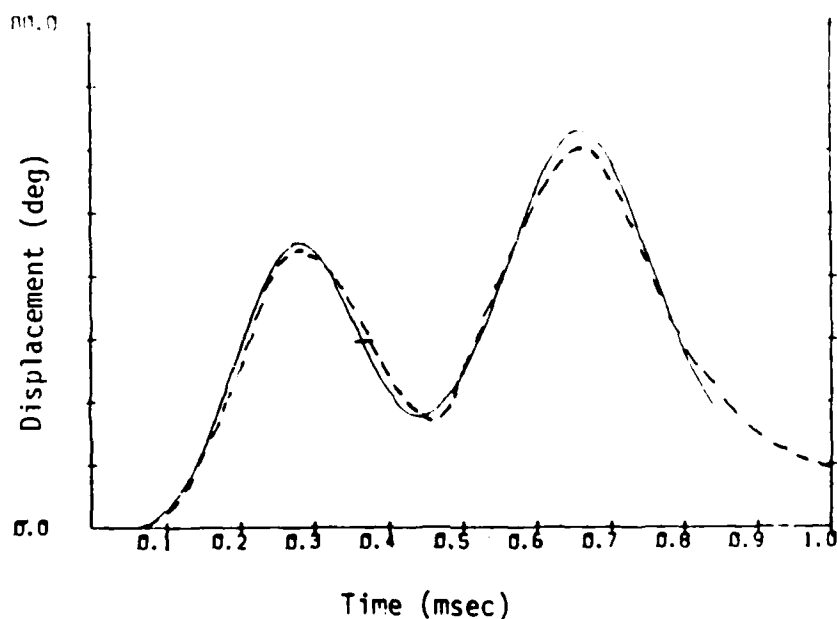

Figure 9. Position-time function of the movement pattern used in the complex spatiotemporal patterning task; the dotted trace is the template, and the solid trace is the subject's response on a given trial.

and size of the errors, providing a basis for a correction on the next trial. The subject can also receive RMS error, computed between the template and the movement, as another kind of KR. Compared to our simpler tasks, this task requires more practice before asymptotes are achieved, and subjects report that they actually enjoy the experiment and find the task challenging.

We have begun to use this task in the ARI work, first in an experiment on relative frequency effects (see our Relative Frequency section presented earlier) as part of Winstein's dissertation. Here, relative frequency is varied by withholding the feedback on various trials, with four relative frequency conditions across two days of acquisition, and various relative frequency conditions in transfer. We have begun to examine the effects of these variables on generalizability, by having subjects transfer to other movement tasks involving different templates. Also, this work asks about relative frequency effects as determiners of subjects' capabilities to detect their own errors, as measured by the accuarcy of subjective estimation conducted in a delayed transfer test. We are excited about this new paradigm, as it seems to solve a number of problems we have had with our simpler motor tasks.

Coincidence-Anticipation Task

A second paradigm is being developed at the same time, but with considerably different methods and goals. Most of the work on feedback and motor learning has involved KR--information about achievement of the movement goal in the environment. It has usually ignored feedback of a potentially more important kind--about the pattern of actions that led to the environmental goal. This feedback is usually termed knowledge of performance (KP; Gentile, 1972), and is in general terms the kind of information that experienced instructors in the military, industry, or sport provide for their learners. Instead of saying simply "You missed the target that time" (KR), the instructor would often say "You jerked the trigger that time" which is, by definition, KP.

Despite its obvious importance for skill training, KP has hardly been studied. In complex skills, feedback about response patterning is difficult to record and quantify, requiring measures of movements that require exhaustive film analyses or other methods, making it generally unsuitable for giving feedback immediately after a trial as is done with KR. KP from experts in real-world skills has not been used, probably because of the subjectivity involved in the detection of errors. A procedure such as used in our spatiotemporal patterning task does not help very much, in our view, as the pattern of action is the goal, and thus KR (about the goal) and KP (about the pattern) are nearly redundant. As a result, we know almost nothing about the role of KP in learning, how it should be scheduled, and what the principles of its operation might be.

Our coincidence-anticipation task, developed in conjunction with a Ph.D. student Douglas Young, attempts to provide a paradigm which will solve these problems. The goal was to develop a task with the following criteria: (a) it should have an environmental goal which is distinct from the pattern of action used to produce it, (b) it should have a movement that allows many variations to

achieve the goal, (c) it should have the movement easily measured by computer methods in the laboratory, and (d) it should represent an analog of some real-life action.

The task is a laboratory analog of hitting a ball moving rectilinearly at constant velocity with a uniplanar batswing perpendicular to the ball's flight path. The moving ball is simulated by a series of LEDs, pulsed so that the lights appear to move toward the subject at a constant velocity. The subject holds a lever, makes a backswing, reverses direction, and produces a forward stroke such that the lever "hits" the moving "ball" at a coincidence point. The task is organized so that the subject can begin the movement at various times, backwing as early or as far as desired, and swing with various patterns of action. This allows for a wide variety of movements which could produce the same outcome (hitting the "ball").

The subject's goal is to maximize a score, analogous to hitting the ball as far as possible, on each trial. Of course, in the real-life case, maximizing the distance a ball will be hit requires that (a) velocity at contact be high, and (b) spatial error at contact be low. So, hitting a real ball involves a trade-off between velocity and accuracy, which we built into the laboratory analog. In our version, the score is the product of velocity at the coincidence point and a weighting coeficient (W),

$$SCORE = VELOCITY \times W,$$

where W is a function of the spatial error, as seen in Figure 10. Here, as the subject makes larger spatial errors, the weight W decreases according to an experimenter-defined schedule. Because large velocity is usually associated with large spatial errors (Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979), in order for the subject to maximize the score, he/she will have to learn the relationship between velocity, spatial error, and score, and produce behavior which is consistent with these principles. This is just as it is in hitting a real baseball, or so it would seem. Subjects are told of this relationship, and the methods for the computer's computation of the score for each trial are described as well.
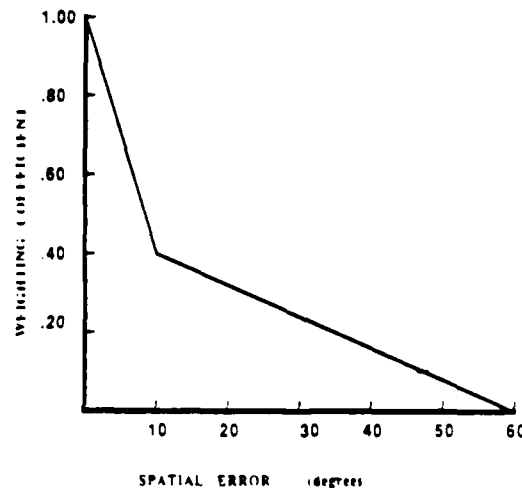

Figure 10. Function for determining the weighting coeficient (W) on a particular trial as determined by the spatial error; as the velocity increases, so usually does the spatial error, thus decreasing W, and reducing the calculated score.

We have conducted two simple experiments to test this paradigm, and to determine its suitability for future work on learning, KR, and KP. In one, the effects of KR (i.e., the score) vs. no-KR in acquisition was studied; and in another the effects of KR vs. KR plus its components (velocity and spatial error) were examined. Generally, learning occurred without KR, but more occurred with KR, and this improvement transferred to the no-KR transfer tests after 10 min and 2 days. The second experiment showed that the additional information about the components provided slight additional benefits over and above KR alone. These results are not surprising, but the experiments were needed in order to be certain that the task was sensitive to feedback manipulations.

But now, the important problem of developing measures of movement patterning is being conducted. Here, we are studying the kinematic measures of movement patterns, such as when the backswing began, how far the backswing went, how consistent the backwing was, when peak acceleration was reached, and so on. We are attempting to determine which aspects of the movement pattern predict success in meeting the movement goal--i.e., in maximizing the score as described earlier. If such features of the movements can be discovered, then we plan to use feedback about them as KP in later experiments. This approach is new and untried, but we are enthusiastic that this paradigm will allow experiments on the role of KP in movement learning, in a way parallel to the work we are now doing on KR. If such principles could be discovered, then they would provide important guidance for instructors in providing movement pattern feedback to their learners, and would certainly be applicable to many Army training settings.

REFERENCES

Adams, J.A. (1986). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, in press.

Armstrong, T.R. (1970). Training for the production of memorized movement patterns. Technical Report No. 26, Human Performance Center, University of Michigan.

Bilodeau, E.A., & Bilodeau, I.M. (1958). Variable frequency of knowledge of results and the learning of simple skill. Journal of Experimental Psychology, 55, 603-612.

Gentile, A.M. (1972). A working model of skill acquisition with application to teaching. Quest, 17, 3-23.

Guthrie, E.R. (1952). The psychology of learning. New York: Harper and Row.

Ho, L., & Shea, J.B. (1978). Effects of relative frequency of knowledge of results on retention of a motor skill. Perceptual and Motor Skills, 46, 859-866.

Hogan, J.C., & Yanowitz, B.A. (1978). The role of verbal estimates of movement error in ballistic skill acquisition. Journal of Motor Behavior, 10, 133-138.

Hull, C.L. (1943). Principles of behavior. New York: Appleton-Century.

Johnson, R.W., Wicks, G.G., & Ben-Sira, D. (1981). Practice in the absence of knowledge of results: motor skill retention. Unpublished manuscript, University of Minnesota.

Lavery, J.J. (1962). Retention of simple motor skills as a function of type of knowledge of results. Canadian Journal of Psychology, 16, 300-311.

Lavery, J.J., & Suddon, F.H. (1962). Retention of simple motor skills as a function of the number of trials by which KR is delayed. Perceptual and Motor Skills, 15, 231-237.

Newell, K.M. (1977). Knowledge of results and motor learning. Exercise and Sport Science Reviews, 4, 195-228.

Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reapraisal. Psychological Bulletin, 95, 355-386.

Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.

Schmidt, R.A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics Press.

Schmidt, R.A., & White, J.L. (1972). Evidence for an error-detection mechanism in motor skills: A test of Adams' closed-loop theory. Journal of Motor Behavior, 4, 143-153.

Schmidt, R.A., & Wrisberg, C.A. (1973). Further tests of Adams' closed-loop theory: Response produced feedback and the error detection mechanism. Journal of Motor Behavior, 5, 155-164.

Schmidt, R.A., Zelaznik, H.N., Hawkins, B., Frank, J.S., & Quinn, J.T. (1979). Motor-output

variability: A theory for the accuracy of rapid motor acts. Psychological Review, 86, 415-451.

Shapiro, D.C. (1977). A preliminary attempt to determine the duration of a motor program. In D.M. Landers & R.W. Christina (Eds.), Psychology of motor behavior and sport (Vol 1). Champaign, IL: Human Kinetics Press.

Shapiro, D.C. (1978). The learning of generalized motor programs. Doctoral dissertation, University of Southern California.

Tolman, E.C. (1932). Purposive behavior in animals and men. New York: Century.

PUBLICATIONS PARTIALLY SUPPORTED BY THIS CONTRACT

Schmidt, R.A. (in press). Controlling the temporal structure of limb movements. The Behavioral and Brain Sciences.

Schmidt, R.A., & Young, D.E. (in press). Transfer of movement control in motor skill learning. In S.M. Cormier & J.D. Hagman (Eds.), Transfer of Learning. Orlando, FL: Academic Press.

Schmidt, R.A. (in press). Toward a better understanding of the acquisition of skill: Theoretical and practical contributions of the task approach. In J.S. Skinner (Ed.), Future directions in exercise and sport research. Champaign, IL: Human Kinetics Press.

Schmidt, R.A. (in press). Toward a better understanding of motor learning. Theoretical and practical contributions of the task approach. Paper to be given at NASPSPA, Tempe, AZ, June, 1986. (Abstract)

Schmidt, R.A. (in press). "Motor" and "action" perspectives on motor behavior. In O. Meijer & K. Roth (Eds.), Complex movement behavior: The motor-action controversy. Amsterdam: North-Holland.

Shapiro, D.C. (in press). An examination of rapid positioning movements with spatiotemporal constraints. Journal of Motor Behavior.

Young, D.E., Schmidt, R.A., Asano, D.L., Shapiro, D.C. (1986). Knowledge of results and complex motor skill learning. Psychology of Motor Behavior and Sport.